Identification of employment concentration areas

Citation for published version (APA):

DOI:
10.1080/09654313.2012.741573

Document status and date:
Published: 01/01/2014

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Identification of Employment Concentration Areas

Frank P. van den Heuvel, Peter W. de Langen, Karel H. van Donselaar, Jan C. Fransoo

School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

Published online: 20 Nov 2012.

To cite this article: Frank P. van den Heuvel, Peter W. de Langen, Karel H. van Donselaar & Jan C. Fransoo, European Planning Studies (2012): Identification of Employment Concentration Areas, European Planning Studies, DOI: 10.1080/09654313.2012.741573

To link to this article: http://dx.doi.org/10.1080/09654313.2012.741573

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Identification of Employment Concentration Areas

FRANK P. VAN DEN HEUVEL, PETER W. DE LANGEN, KAREL H. VAN DONSELAAR & JAN C. FRANSOO

School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

(Received July 2012; accepted September 2012)

ABSTRACT This article presents a method to identify “Absolute and Relative Employment Concentration (AREC) areas” for a particular industry. Two novel characteristics of the method are that it simultaneously analyses AREC, and that it combines spatial concentration per area with the spatial concentration in neighbouring areas. The method is easy to understand and apply. It is developed to assist regional policy makers and corporate decision-makers with their investment decisions related to new infrastructure or plants. The identification of concentration areas also allows for analysing the performance of these areas in relation to characteristics such as infrastructure availability and the housing and labour market. This can yield new academic insights that are relevant for regional planners. An application of the newly developed method to five industries in a Dutch province subdivided into 502 areas illustrates the value of the method in comparison to other methods.

Introduction

In recent years, there has been a lot of attention for the concept of spatial clustering (Gordon & McCann, 2000; Porter, 2000). Research in new economic geography (Fujita & Krugman, 2004) explains spatial concentration of industrial activity by agglomeration economies. These economies are related to proximity and distance; the closer firms are together, geographically and organizationally (Torre & Rallet, 2005), the more synergies between these firms play a role. Agglomeration economies have been researched extensively (e.g. by Krugman, 1991; Malmberg & Maskell, 2002), always (partly) based on the three major sources of localization economies (Marshall, 1956): labour market pooling, inputs sharing and knowledge spillovers.

Regional, national and supra-national policy makers all embrace the cluster concept in regional development policies. At the European level, several reports deal with clusters.
related to different sectors (European Commission, 2002, 2008, 2011), at national level, the in 2010 installed Dutch government aimed to encourage the growth of regional clusters, and at the level of the Dutch provinces, governments use the concept of spatial clustering to stimulate regional development. Although governments often use the concept of clustering, none of them describes what is meant precisely, and probably many of them mean different things (Lundequist & Power, 2002; Martin & Sunley, 2003; Newlands, 2003). This article presents a new method to identify the areas where a particular industry is spatially concentrated. This method is especially relevant for analysing spatial clustering at a small geographic scale.

Porter (1998, p. 78) defines clusters as “geographic concentrations of interconnected companies and institutions in a particular field”. Although this definition is intuitively clear, several components of it can be interpreted differently in different contexts. First, the concept of geographic concentration is researched extensively (see, e.g. Bickenbach and Bode (2008) and Fratesi (2008) for general overviews of spatial concentration measures) and has many different interpretations. Second, the interconnection between firms can be interpreted in many different ways (e.g. buyer–supplier relationships, but also competition for the same type of regional labour, and so on). Third, the way Porter’s “particular field” is defined is often unclear. This article focuses on the first aspect: geographical concentration. It develops a method to identify those areas in a particular region in which a particular sector (or set of related sectors) is spatially concentrated. Although Benneworth et al. (2003) legitimately remark that the presence of one cluster element, like, for example, spatial concentration, does not mean that other cluster elements, like, for example, cooperation, are also present, spatial concentration is mostly seen as a prerequisite for clusters. An analysis into transaction links between companies located in these spatial concentration areas (Hanssens et al., 2012) can complement this analysis, but requires different data and, hence, is beyond the scope of this article.

Many techniques have been developed to measure spatial concentration of firms or employment. Most of the spatial concentration measures are defined based on rather large geographical areas, such as economic areas or states in the US and provinces in Europe, while spatial concentration is also observable at lower geographical scales (Arauzo-Carod & Viladecans-Marsal, 2009; Van Soest et al., 2006). For instance, input sharing in the chemicals industry often leads to co-location of chemical firms at the same site and in logistics, co-location leads to reduced transport costs (Taniguchi et al., 1999). Despite the relevance of co-location in small geographic areas, limited attention has been given to the issue of identifying small-scale areas with concentrations of specific industries. This article addresses this issue. The method is applied to identify spatial concentration in 502 areas in one Dutch province. Hence, the spatial aggregation level is low; the identified concentration areas have a surface of 10–30 km². These areas are termed Absolute and Relative Employment Concentration (AREC) areas, for a particular industry within a specific region. As suggested by the name, the method identifies areas with both relative and absolute concentrations. The term “AREC area” is preferred over the term “cluster”, as the issue of interconnectedness is not addressed. Also the term “hotspot” is not used, even though this term is often used for areas with large concentrations of activities in specific industries. However, the term is also often used with a different meaning, namely the concentration of innovative activity (Pouder & St. John, 1996).
The proposed method contains two steps. In step 1, the spatial concentration of industries in the region is measured. This analysis is relevant, since in the absence of spatial concentration, economic activities are spread out evenly in different areas and no AREC areas can be identified. This is unlikely to be the case in most industries (Guillain & Le Gallo, 2010), but this analysis also shows the order of magnitude of spatial concentration in an industry, which is relevant for the interpretation of the AREC areas. In step 2 of the method, AREC areas are identified, based on both absolute and relative concentrations. Existing spatial concentration methods focus on measuring relative spatial concentration (an above average share of an industry in a region). This is explained by the goal of these measures to identify differences in concentration between different industries, different regions and/or over time. For that purpose, absolute concentration (large numbers of employment/firms in an industry in a region) is not relevant. Some literature uses absolute spatial concentration (Wennberg & Lindqvist, 2010), and although Jing and Cai (2010) argue that using relative or absolute concentration results in different conclusions, no literature was found in which absolute and relative spatial concentration are considered simultaneously.

The identification of AREC areas is especially valuable for regional policy makers, who can make more informed decisions about land use and infrastructure, and corporate decision-makers, who can make better informed location decisions. Both decision-makers can benefit from more insight in concentration of a specific industry in a specific area, both for large (e.g. to focus on the development of a small number of clusters) and small (e.g. to decide where to invest in industrial parks and infrastructure) geographical areas. This article deals with small geographical areas. As these small geographical areas (e.g. postal code areas) generally differ in employment size, only looking at the absolute concentration of the industry’s employment is not enough. In this approach, areas with a high overall employment level (e.g. urban areas) are often selected as concentration areas, even when the specific industry is not overrepresented. Relative concentration of an area’s employment in an industry is not enough either, as using only relative concentration results in the selection of areas with a very low general employment size, when a few firms (or even just one) in a specific industry are located there. This is especially relevant for the low aggregation level addressed in this article. As an extreme example: a shepherd with cattle in the desert does not provide a basis for promoting a dairy cluster there, nor is relevant for cattle farmers in search for an investment location. Hence, concentration areas of a particular industry are best identified based on both absolute and relative concentrations.

AREC areas in a particular region are relevant constructs for research. With regard to infrastructure availability, it can be tested whether or not the location of (logistics) AREC areas is related to intermodal facilities, distances to the highway network or both. In various industries, the relation between AREC areas and specialized education institutions can be analysed. The relation between AREC areas for high-value industries and housing prices can be studied. Furthermore, it can be tested whether land or industry specific property values are higher in AREC areas than outside these areas. Finally, it can be tested, for different industries or in general, whether or not employment growth in AREC areas is different from employment growth in other areas.

The remainder of this article is organized as follows. The second section will start with a short overview of the literature on spatial concentration measures. The third section presents the newly developed method to identify AREC areas. Then, the method is applied to
Literature on the Measurement of Spatial Concentration

Many spatial concentration measures have been developed. Extensive overviews of different measures are given by Bickenbach and Bode (2008) and Fratesi (2008). All measures deal with relative concentration, as absolute concentration simply involves an employment count and cannot be compared in a meaningful way across areas (that differ in size). This section gives a short overview of the most important measures.

For the brief overview in this article, it is useful to make a distinction between two categories of measures. The first category consists of measures for which the total region under study is divided into \( K \) areas and the spatial concentration is analysed per predefined area. The measures in this category ignore the spatial relationship, i.e. the distance and shared borders, between the different areas. For this reason, these measures are called a-spatial measures (although these still are spatial concentration measures). In the second category of measures, analogously called spatial measures, the region is not divided in predefined areas or the spatial relationship between the predefined areas is explicitly analysed. Well-known and commonly used a-spatial measures are the locational Gini coefficient (Krugman, 1991), the EG-index (Ellison & Glaeser, 1997), the MS-index (Maurel & Sedillot, 1999) and the D-index (Mori et al., 2005). The most commonly used spatial measures were developed by Moran (1950) and Duranton and Overman (2005).

All a-spatial measures share two well-known problems (Bickenbach & Bode, 2008), namely the modifiable areal unit problem (MAUP) and the checkerboard problem. Both result from the division of the total region into predefined areas. The MAUP arises from dividing heterogeneous continuous space into areas (Arbia, 2001a): a group of firms that are located close together can be grouped into one area or spread over several areas, depending on the defined borders. This influences the degree of spatial concentration of these firms. The checkerboard problem concerns neglecting relevant information on the locations of or distances between regions (Arbia, 2001b): the a-spatial measures do not make a distinction between two areas with relatively much employment that are geographically close together or far apart.

Several spatial (or spatial association) measures have been developed to cope with these problems. These measures can be broadly subdivided into distance-based measures and neighbourhood measures (Anselin, 1996). For distance-based measures, the region is not subdivided into different areas. The intuitive idea behind the distance-based approaches is that when an industry is concentrated in one or more parts of the region, firms are located at a shorter distance from each other then when they are randomly located over space. Distance-based measures were developed by, among others, Duranton and Overman (2005) and Marcon and Puech (2010). Since these measures are based on the distances between the industry’s firms, data on the exact locations of the firms are required. To calculate Euclidean distances, the geographical coordinates of all firms in the data set are needed, and to calculate distances via transport networks, both the exact addresses of the firms, transport network structures, and software to analyse these have to be available. A major disadvantage of these measures is that this is often not the case. Hence, the method presented in this article will not make use of distance-based approaches.
Instead, one of the a-spatial measures is used to measure relative spatial concentration in the first step of the method. The second step of the method uses parts of the theory of the (local) Moran’s $I$ (Cliff & Ord, 1973) to minimize the effect of subdividing the region into predefined areas. Moran’s $I$ is one of the most important neighbouring methods. Although the method does not completely solve the MAUP nor the checkerboard problem, for the purposes in this article (to develop a method that is easy to apply and understand), it is the best alternative.

The remainder of this section describes the important characteristics of the a-spatial measures and evaluates their relevance for the AREC area identification method. The measures will be presented to evaluate spatial concentration of industry $i \in \{1, \ldots, I\}$, with $I$ all relevant industries in the total region that can be subdivided into $K$ different areas $k \in \{1, \ldots, K\}$.

One of the oldest contributions to the literature of measuring spatial concentration is by Krugman (1991), who developed the locational Gini coefficient. This index makes use of the location quotient (LQ): $LQ_{i,k} = \frac{s_{i,k}}{s_k}$, being a measure to analyse relative spatial concentration, defined as an area $k$’s share of industry $i$’s employment ($s_{i,k}$) divided by that area’s share of total employment ($s_k = \sum_{i=1}^{I} s_{i,k}$). The locational Gini coefficient is widely used due to its ease of computation and its limited data requirements (Bertinelli & Decrop, 2005). However, the locational Gini coefficient does not control for industrial concentration (i.e. concentration of employment in a small number of large firms).

Alternative measures to analyse relative spatial concentration are, among others, $\gamma_{EG}$ (Ellison & Glaeser, 1997), $\gamma_{MS}$ (Maurel & Sédillot, 1999) and the D-index (Mori et al., 2005). The $\gamma_{EG}$ and the $\gamma_{MS}$ have the same properties: the indices are scaled such that these control for industrial concentration. Moreover, the indices are designed to enable the comparison of the concentration of different industries, different regions and/or over time. The major drawback of these indices is that the outcomes are hard to interpret; for both indices, boundaries of 0.02 and 0.05 are used to define regions of no concentration ($\gamma < 0.02$), intermediate concentration ($0.02 \leq \gamma \leq 0.05$) and high concentration ($\gamma > 0.05$). However, these boundaries are arbitrary (Duranton & Overman, 2005).

The D-index (Mori et al., 2005) is designed to be statistically testable, since none of the above-described measures are. However, this index can only be used for an analysis of spatial concentration of firms, because it is based on the independence of the single units, which means that the D-index is not suitable for the measurement of employment concentration (Fratesi, 2008).

The AREC Area Identification Method

This section describes the AREC area identification method. Step 1 analyses whether an industry’s employment is spatially concentrated in the first place. Step 2 of the method uses three consecutive steps to identify AREC areas. This section explains the method that is visualized in Figure 1.

**Step 1: The Degree of Spatial Concentration**

Step 1 of the method is based on the locational Gini coefficient (Krugman, 1991), because this index is easiest to understand and can be interpreted intuitively. The commonly mentioned drawback is that it does not compensate for industrial concentration. However, for
step 1, this is not a problem, because the sole purpose of the analysis is to get a rough idea about whether or not there is some degree of spatial concentration, since in the absence of spatial concentration, identifying AREC areas is meaningless. In addition, the locational Gini coefficient uses the LQ per area, which can directly be used in step 2 of the method as a measure for relative spatial concentration of industry $i$ in a specific area $k$. When either $\gamma_{EG}$ or $\gamma_{MS}$ would be used, it is much harder to translate this aggregate measure to a local measure per area, due to the compensation for industrial concentration on the aggregate level. In this article, the following definition of the locational Gini coefficient is used (Guillain & Le Gallo, 2010; Kim et al., 2000):

$$G_{LQ,i} = \frac{\Delta_{LQ,i}}{4LQ_i}$$

$$\Delta_{LQ,i} = \frac{1}{K(K-1)} \sum_{k=1}^{K} \sum_{j=1}^{K} |LQ_{i,k} - LQ_{i,j}|$$

with $LQ_i$ the average LQ of industry $i$ over all areas $K$. The locational Gini coefficient is equal to zero when no relative spatial concentration is measured, meaning that the spatial distribution of the industry’s employment is equal to the spatial distribution of employment in general ($LQ_{i,k} = 1$ for all $k$). The maximum value of this index is equal to 0.5, when the industry’s employment is completely concentrated in one area.

The locational Gini coefficient shows whether an industry’s employment is spatially concentrated. For industries where this is not the case, there is no value to identify AREC areas, since employment is evenly spread over the region.

**Step 2: The Identification of AREC Areas**

After it is concluded that employment in an industry is spatially concentrated (step 1 of the method), step 2 identifies where this concentration takes place. The method is novel in two
Identification of Employment Concentration Areas

ways. First, AREC areas will be identified based on absolute and relative concentration simultaneously. Second, not only spatial concentration per area will be used to identify AREC areas, but also spatial concentration in neighbouring areas will be included. As Hallencreutz and Lundequist (2003) legitimately conclude, clusters (or spatial concentration areas) seldom confirm to boundaries of administrative areas. Hence, the method should also consider neighbouring administrative areas that together can be identified as AREC areas. For this, neighbours have to be defined for all areas in the region. This can be done based on a shared border between particular predefined areas or on the distance between the centre-points of these areas. This neighbourhood structure is formalized in a spatial weight matrix $W$, with elements $w_{jk} > 0$ if areas $j$ and $k$ are neighbours and $w_{jk} = 0$ if these areas are not neighbours or if $j = k$. This matrix has to be determined before the method described below is applied.

A stepwise method is developed to determine where a particular industry concentrates. Step 2 of the method itself consists of three steps:

2.1. Remove areas that do not have the potential to become (part of) an AREC area, since either the relative or the absolute concentration is too low.

(a) Define for both the absolute employment values and the LQ values a cut-off value $e_{\text{min},1}$ and $\text{LQ}_{\text{min},1}$, respectively.

(b) Code all areas $k \in \{1, \ldots, K\}$ based on their absolute employment level $e_{i,k}$ in industry $i$ and their $\text{LQ}_{i,k}$ as compared with these cut-off values: $b_{i,k} = 1$ if $e_{i,k} \geq e_{\text{min},1} \cap \text{LQ}_{i,k} \geq \text{LQ}_{\text{min},1}$ and $b_{i,k} = 0$ otherwise. Remove all areas $k$ with $b_{i,k} = 0$.

2.2. In the set of remaining areas, combine areas that are neighbours.

(a) Determine $B_{i,j,k}$ joins between two neighbouring areas $j$ and $k$, with both $b_{i,j} = 1$ and $b_{i,k} = 1$: $B_{i,j,k} = w_{jk}b_{i,j}b_{i,k}$, with $w_{jk} = 1$ if areas $j$ and $k$ are neighbours and $w_{jk} = 0$ otherwise. Furthermore, $w_{jk} = 0$ if $j = k$.

(b) For all $j$ and $k$ for which $B_{i,j,k} = 1$, define a new area $m \in \{K + 1, \ldots, M\}$, with employment being the sum of the employment values of the different predefined areas: $e_{i,m} = e_{i,j} + e_{i,k}$. Area $m$ can exist of more than two predefined areas: for example, if $B_{i,j,l} = 1 \cap B_{i,j,l} = 1$, then $e_{i,m} = e_{i,j} + e_{i,k} + e_{i,l}$, etc. In addition, calculate $\text{LQ}_{i,m}$. The total number of newly created areas is equal to $(M - K)$.

2.3. Identify AREC areas based on absolute and relative employment levels.

(a) Define for both the absolute employment values and the LQ values a second cut-off value: $e_{\text{min},2} \geq e_{\text{min},1}$ and $\text{LQ}_{\text{min},2} \geq \text{LQ}_{\text{min},1}$.

(b) Determine $\beta_{i,c}$ for all areas $c \in \{1, \ldots, M\}$: $\beta_{i,c} = 1$ if $e_{i,c} \geq e_{\text{min},2} \cap \text{LQ}_{i,c} \geq \text{LQ}_{\text{min},2}$, and $\beta_{i,c} = 0$ otherwise. All areas $c$ with $\beta_{i,c} = 1$ are identified as AREC areas of industry $i$. If both $\beta_{i,k} = 1$ and $\beta_{i,m} = 1$, with $e_{i,m} = e_{i,k} + e_{i,j}$ for any $j \in \{1, \ldots, M\}$, only the larger area $m$ is defined as an AREC area.

Notice that the method combines absolute and relative spatial concentration, by considering both absolute employment per area and the LQ. Both are used to determine whether an administrative area is (part of) an AREC area for that industry $i$. Furthermore, in step 2.2, neighbouring areas with relatively high AREC for industry $i$ are combined, to check whether the AREC area actually consists of two or more administrative areas. For this, the local Moran’s $I$ (Anselin, 1995; Guillaun & Le Gallo, 2010) could have been used, if only either absolute employment concentration or relative employment concentration was analysed, since this index can only be calculated on one variable (Cliff & Ord, 1973) and
no clear procedure exists to combine these (absolute and relative) values into one variable. Nevertheless, a binary variable $b_{ik}$ can be created, which indicates whether a particular area has the potential to become (part of) an AREC area. In step 2.2, area $k$’s value of this binary variable is compared with the value of one of its neighbours, area $j$, defined based on the predefined spatial weight matrix $W$. When the variable is equal to one for both of these two neighbouring areas, it indicates that the potential AREC area may be larger than one administrative area. Hence, a new area $m$ is created, containing both the administrative areas $j$ and $k$. By combining two or more of the administrative areas, the method (partly) compensates for the MAUP and checkerboard problem. In step 2.3, all areas, now consisting of one or more administrative areas, are evaluated against two new cut-off values. These values can be equal to the values defined in step 2.1. However, the cut-off values defined in step 2.1 only determine which administrative areas are immediately removed from the analysis, because these have really low (absolute or relative) employment concentration in industry $i$. It is possible that two administrative areas that by themselves are no AREC areas can be a combined AREC area. Thus, when combined areas are created in step 2.2, only the areas that would not even have potential to become (part of) an AREC area have to be removed. Later on in step 2.3, where AREC areas are identified, higher cut-off values are more useful, since then only the areas with very high (absolute and relative) employment concentration in industry $i$ are of interest. Hence, it is advised to use higher cut-off values in step 2.3 than in step 2.1 of the method.

An important part of this method is the determination of the cut-off values $e_{min,1}$, $e_{min,2}$, $LQ_{min,1}$ and $LQ_{min,2}$. For the cut-off values used in step 2.1, it is proposed to use conservative values, as the method is not iterative, meaning that removed areas will not be analysed. For the LQs, areas with an LQ larger than or equal to 1 are selected, since these areas are commonly stated to be, at least a bit, specialized in that industry $i$ (Guillain & Le Gallo, 2010). For the absolute value, the commonly used cut-off value is the average employment over all areas, also used in the (local) Moran’s $I$. Instead of these cut-off values, O’Donoghue and Gleave (2004) propose to convert LQ values to standardized LQ values and use statistical significance to determine which LQ values are extremely high. However, in this article, cut-off values will be used, since these can be set by the users of the method directly. The goal of the method is to identify the areas in which a particular industry is spatially concentrated, with cut-off values depending on the goal of the user of the method, and not to analyse whether and where this spatial concentration is significant.

The cut-off values used in step 2.3 can be chosen differently for different purposes. Consider again the two situations described in the beginning of the article. In the example of the policy maker, a low number of AREC areas may be preferred and hence, the cut-off values are chosen relatively high. In the other example, in which AREC areas are identified to assist a location decision for an individual company, more AREC areas may be desirable, since a manager may want to have a complete overview of relevant locations, and the cut-off values can be somewhat lower. Furthermore, in this article it is argued that both absolute and relative spatial employment concentration are needed to define these areas, but not whether these are equally important. This may also differ per decision maker; for some reason, some may want to post stricter cut-off values on the absolute employment concentration than on the relative employment concentration or vice versa. In the next section, rather high cut-off values are chosen (in line with the policy perspective for both the absolute and relative concentrations to identify AREC areas in one Dutch province.

F.P. van den Heuvel, P.W. de Langen, K.H. van Donselaar & J.C. Fransoo
Application of the AREC Area Identification Method

This section applies the AREC area identification method to five different industries in North Brabant, one of the southern provinces (NUTS 2) of the Netherlands. Figure 2 shows the location of the study region. First, the data set used for the analysis is discussed and afterwards the results of the analysis are presented.

Material Used

The method is applied to five industries with a different extent of spatial concentration and of a different size. The industries used in this study are manufacturers of chemicals and chemical products (chemical production), research and development (R&D), logistics, retail trade and construction. Appendix 1 presents a list of the industry codes used for defining these industries. In line with the findings of Guillain and Le Gallo (2010), two of these industries tend to concentrate spatially in North Brabant (chemical production and R&D) and two others tend to distribute equally over space (construction and retail trade). Logistics is included since it is a relatively large industry in North Brabant (like construction and retail trade), but still has a relatively high spatial concentration. Furthermore, this industry is especially relevant from a policy perspective, since logistics activities generate relatively many transport flows and occupy relatively much space.

Figure 2. NUTS 2 areas in Western Europe; North Brabant is the highlighted area.
For the analysis, the database containing all business establishments in North Brabant in 2009 is used. After a standard clean-up, this database contains 160,647 establishments in total. Furthermore, governmental organizations and establishments related to education, health services, culture and recreation activities were deleted from the database. The remaining database covers 129,555 establishments with a total employment of 856,668, which gives an average number of employees per establishments of 6.61, with a standard deviation of 41.64 and a median of 1. This total database is used to determine the LQ per area for the five above-mentioned industries.

The total region is subdivided in 502 four-digit postal code areas. Table 1 presents descriptive statistics of the spread of employment over the 502 different postal code areas. Since some of the listed industries are only present in part of all four-digit postal-code areas (e.g. the chemical industry only has establishments in 102 of the 502 four-digit postal code areas), averages without the areas with no industry employment are also calculated (see the fourth column of Table 1). These are used in the analysis later on.

### Step 1: The Degree of Spatial Concentration

Table 2 presents the locational Gini coefficients. The chemical and R&D industries are highly spatially concentrated, the retail trade and construction industries are less concentrated, and logistics is somewhere in between. These outcomes are comparable to what Guillain and Le Gallo (2010) found based on communes in and around Paris, as can be seen in Table 3. French communes have an average surface of 15 km², while the four-digit postal code areas in North Brabant have an average surface of 10 km². Furthermore,

---

**Table 1.** Descriptive statistics of the number of employees per four-digit postal code area in North Brabant

<table>
<thead>
<tr>
<th></th>
<th>Sum</th>
<th>Average</th>
<th>Average if not 0</th>
<th>Standard deviation</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>856,668</td>
<td>1706.51</td>
<td>1713.34</td>
<td>2241.62</td>
<td>888.5</td>
<td>17,353</td>
</tr>
<tr>
<td>Chemical production</td>
<td>8833</td>
<td>17.60</td>
<td>86.60</td>
<td>112.71</td>
<td>0</td>
<td>1821</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>7096</td>
<td>14.14</td>
<td>36.96</td>
<td>192.29</td>
<td>0</td>
<td>4258</td>
</tr>
<tr>
<td>Logistics</td>
<td>90,388</td>
<td>180.06</td>
<td>221.54</td>
<td>398.70</td>
<td>30.5</td>
<td>2975</td>
</tr>
<tr>
<td>Retail trade</td>
<td>118,513</td>
<td>236.08</td>
<td>244.36</td>
<td>375.19</td>
<td>110</td>
<td>3436</td>
</tr>
<tr>
<td>Construction</td>
<td>85,646</td>
<td>170.61</td>
<td>173.37</td>
<td>212.46</td>
<td>91</td>
<td>2032</td>
</tr>
</tbody>
</table>

**Table 2.** Locational Gini coefficients based on the employment in North Brabant

<table>
<thead>
<tr>
<th>Industry</th>
<th>Gini (based on LQ values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical production</td>
<td>0.4750</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>0.4620</td>
</tr>
<tr>
<td>Logistics</td>
<td>0.2984</td>
</tr>
<tr>
<td>Retail trade</td>
<td>0.2224</td>
</tr>
<tr>
<td>Construction</td>
<td>0.1910</td>
</tr>
</tbody>
</table>
both study regions contain urban areas as well as rural areas, and hence, the studies can be compared.

**Step 2: The Identification of AREC Areas**

Based on the locational Gini coefficients found for the five different industries, only AREC areas for the chemical, R&D and logistics industry will be identified. An analysis into the AREC areas for the retail industry and the construction industry is not relevant, since there hardly is any spatial concentration in these two industries; for example, for the construction industry, the AREC areas contain one area that consists of 34 four-digit postal code areas, which cannot be called a concentration area anymore on this low spatial aggregation level.

In line with the third section, in step 2.1 of the method, the average employment per area is used as a cut-off value for the absolute employment level and 1 as a cut-off value for the LQ. The cut-off values used in step 2.3 depend on the purpose of the analysis. In this article, the policy perspective is chosen and hence, rather high cut-off values are used in step 2.3 of the method. For the absolute employment level, the maximum of the 90th percentile and the bound defined in step 2.1 is used as a cut-off value, meaning that only the top 10% areas can qualify as AREC areas. For the LQ, the minimum of the 90th percentile and 2 is determined; the LQ cut-off value is equal to the maximum of this value and the cut-off value defined in step 2.1 of the method. As previously argued, all statistics used as bounds are calculated based on areas in which the employment of the industry is larger than zero. Table 4 presents the cut-off values used for the industries under study. As spatial weight matrix $W$, a matrix based on the borders shared by the areas is used: for every row-area $j$, it is determined whether it shares a border with a column-area $k$; if this is the case $w_{jk} = 1$, and $w_{jk} = 0$ otherwise.

The method is shown graphically for the chemical industry in Figure 3. The figure presents all four-digit postal code areas based on their absolute employment in chemical

### Table 3. Locational Gini coefficients based on employment in and around Paris (Guillain & Le Gallo, 2010)

<table>
<thead>
<tr>
<th>Industry</th>
<th>Gini (based on LQ values)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical production</td>
<td>0.4752</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>0.4922</td>
</tr>
<tr>
<td>Wholesale trade</td>
<td>0.3898</td>
</tr>
<tr>
<td>Transportation and communication</td>
<td>0.3797</td>
</tr>
<tr>
<td>Consumer services</td>
<td>0.3220</td>
</tr>
<tr>
<td>Construction</td>
<td>0.2972</td>
</tr>
</tbody>
</table>

### Table 4. Cut-off values used for the application of the AREC area identification method

<table>
<thead>
<tr>
<th>Industry</th>
<th>$e_{\text{min},1}$</th>
<th>$\text{LQ}_{\text{min},1}$</th>
<th>$e_{\text{min},2}$</th>
<th>$\text{LQ}_{\text{min},2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical production</td>
<td>86.60</td>
<td>1</td>
<td>193.70</td>
<td>2</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>36.96</td>
<td>1</td>
<td>36.96</td>
<td>1.90</td>
</tr>
<tr>
<td>Logistics</td>
<td>221.54</td>
<td>1</td>
<td>630.90</td>
<td>2</td>
</tr>
</tbody>
</table>
production (X-axis) and their LQ based on the employment in chemical production (Y-axis). All open diamonds represent one of the 502 postal code areas. In step 2.1, all areas to the left and below the line based on the first cut-off values are removed from the analysis; the areas to the right and above this line are considered in step 2.2. The combined areas created in step 2.2 are characterized by squares in the figure. Finally, the AREC areas defined in step 2.3 are pictured as filled diamonds, based on the second cut-off values.

Figure 4 presents the AREC areas of the chemical production industry geographically. In this figure, postal code areas are coloured red if these are AREC areas by themselves ($e_{1,k} \geq e_{\text{min},2} \cap LQ_{1,k} \geq LQ_{\text{min},2}$ for $k \in \{1, \ldots, 502\}$ and are coloured orange if these are part of a combined AREC area. The 10 AREC areas consist of 12 of the 502 four-digit postal code areas (3%) and together account for 74% of all employment in chemical production in North Brabant.

The same analysis is conducted for the R&D industry and the logistics industry (geographical representations for these industries can be found in Figures A2(c) and 5(c), respectively). The 10 AREC areas for R&D account for 83% of all R&D employment in North Brabant; the 19 AREC areas for logistics account for 53% of all logistics employment in the province. These numbers correspond with the locational Gini coefficients, which indicated that both the chemical production industry and the R&D industry are more spatially concentrated than the logistics industry.

**Comparison with Other Methods**

This section compares the identified AREC areas with the areas that would have been identified as concentration areas when only absolute or relative spatial concentration
was used as a criterion. Tables 5 and 6 demonstrate that identifying AREC areas with the method described in the third section results in different areas than using either one of these concentration variables alone. The tables present the number of postal code areas that are selected based on the AREC area identification method and the areas that would have been selected if only absolute concentration (Table 5) or relative concentration (Table 6) would have been used. For example, from Table 5 it becomes clear that if only absolute concentration was used as a criterion to identify concentration areas for logistics, 41 four-digit postal code areas would have been identified as concentration areas. Only 29

![Figure 4. Geographical representation of the AREC areas for chemical production.](image)

<table>
<thead>
<tr>
<th>Table 5. Number of postal code areas identified as AREC areas and the number of postal code areas with only high absolute employment concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_{i,k} \geq e_{\text{min},2}$</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td><strong>Chemical production</strong></td>
</tr>
<tr>
<td>AREC</td>
</tr>
<tr>
<td>Non-AREC</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td><strong>R&amp;D</strong></td>
</tr>
<tr>
<td>AREC</td>
</tr>
<tr>
<td>Non-AREC</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td><strong>Logistics</strong></td>
</tr>
<tr>
<td>AREC</td>
</tr>
<tr>
<td>Non-AREC</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
of these postal code areas are also identified as AREC areas using the newly developed method. Hence, only using absolute concentration for logistics would have resulted in identifying 12 postal code areas as concentration areas, while in these areas the relative concentration (measured using the LQ) of logistics employment is relatively low.4

Tables 5 and 6 show the benefits of the AREC area identification method. First, based on the comparison made for the chemical sector, it can be seen that identifying concentration areas based on relative concentration is not satisfying: Table 6 shows that there are 11 postal code areas not being labelled as AREC areas that have an LQ larger than the second cut-off value. The non-AREC areas with an LQ larger than the cut-off value have really low absolute employment levels, both in the chemical industry and in general. Hence, these indeed are the “shepherd in the desert”—areas: the lowest industry employment found for these areas is equal to 1, but with a total employment of 21 the LQ still is equal to 4.62, while it is by far an interesting area for policy or corporate decision-makers. This effect can also be seen with the other two industries analysed.

Second, the tables show that absolute employment concentration alone is also not enough. This is best shown by using the R&D industry as an example. Table 1 showed for this industry a maximum number of employees per postal code area of 4258, with an average of only 37. Actually, there is only one postal code area with really large industry employment, while the 15 remaining areas with $e_{i,k} \geq e_{\text{min},2}$ (which is above the 90th percentile of the absolute employment level) score relatively close to the average. To identify more than one concentration areas for R&D, the LQ becomes important. Table 5 shows that there are six postal code areas with $e_{i,k} \geq e_{\text{min},2}$ that are not identified as AREC areas, because the LQ of R&D in these areas is too low. Policy and corporate decision-makers would not identify these areas as important for R&D.

Third, the tables show the value of combining neighbouring areas. This can best be seen in the logistics sector. For this sector, Table 5 shows that there are 13 postal code areas that are identified as AREC areas while $e_{i,k} < e_{\text{min},2}$ for these areas, because these are combined with other postal code areas. Similarly, Table 6 shows that there are 13 (different) postal code areas that are identified as AREC areas while $LQ_{i,k} < LQ_{\text{min},2}$ for these areas. This shows the value of combining the areas: if this was not done, these areas would not

<table>
<thead>
<tr>
<th>Chemical production</th>
<th>$LQ_{i,k} \geq LQ_{\text{min},2}$</th>
<th>$LQ_{i,k} &lt; LQ_{\text{min},2}$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREC</td>
<td>12</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Non-AREC</td>
<td>11</td>
<td>479</td>
<td>490</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>479</td>
<td>502</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>10</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Non-AREC</td>
<td>10</td>
<td>482</td>
<td>492</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>482</td>
<td>502</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Logistics</th>
<th>$LQ_{i,k} \geq LQ_{\text{min},2}$</th>
<th>$LQ_{i,k} &lt; LQ_{\text{min},2}$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREC</td>
<td>29</td>
<td>13</td>
<td>42</td>
</tr>
<tr>
<td>Non-AREC</td>
<td>18</td>
<td>442</td>
<td>460</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>455</td>
<td>502</td>
</tr>
</tbody>
</table>
Identification of Employment Concentration Areas

Figure 5. (a) Absolute logistics employment concentration. (b) Relative logistics employment concentration (LQ). (c) AREC areas for logistics.
have been identified as concentration areas, while this only is due to the rather arbitrary division of the total region into administrative areas.

Since especially the logistics industry shows large differences between the AREC areas and the areas identified based on only one concentration variable, Figure 5 shows these areas geographically. Figure 5(a) presents the areas based on absolute employment concentration only, Figure 5(b) presents the areas based on relative employment concentration only, and 5(c) presents the AREC areas for logistics. Many postal code areas with an LQ above 2 are not identified as AREC areas for logistics (because they essentially are “shepherds in the desert”); for example, the eastern part of the province contains seven such areas and no AREC area. In addition, these maps show another advantage of the AREC area identification method over the use of only one concentration variable. Figure 5(a) presents 29 concentration areas (possibly consisting of multiple postal code areas), Figure 5(b) even presents 34 concentration areas, while 5(c) only presents 19 AREC areas. Hence, using the AREC area identification method, decision-makers only have to choose between a smaller number of areas. Furthermore, based on the AREC area identification method, larger combined areas are identified, while the concentration areas identified based on either absolute concentration or relative concentration are very fragmented. The other industries analysed in this article show similar patterns, as shown in Appendix 2.

Conclusions

In this article, a method is developed to identify AREC areas. The identification of these areas is relevant for various purposes: (1) for evaluating whether the agglomeration
economies described by Marshall (1956), Krugman (1991) and Porter (2000) are also relevant on a low geographical aggregation level, (2) for analysing relationships between the location of these AREC areas and variables such as infrastructure availability, distances to specialized higher education institutions, land value and housing prices, and (3) for analysing spatial employment development over time. Since the method is very easy to understand and apply, it is a valuable tool for policy makers as well as corporate decision-makers. Policy makers, for example, can use this method to identify the important areas to invest in and substantiate the choice for these areas rather easily, by means of the rather simple graphs constructed based on the method.

As compared with already existing spatial concentration measures, this method has two novel characteristics. First, existing measures only consider relative spatial employment concentration, while the newly developed method combines this with absolute spatial employment concentration. A comparison of AREC areas and areas with either absolute or relative employment concentration shows that combining absolute and relative employment levels simultaneously filters out areas that are less relevant both from a policy and a location decision perspective. This especially is relevant on a low geographical aggregation level, as used in this article. Second, AREC areas can consist of more administrative areas, which reduce the well-known problems of subdividing the total region into administrative areas. In this way, most of the commonly known problems related to a-spatial measures are dealt with. The advantages of the developed method were demonstrated by applying the method to five different industries in a Dutch province.

Although the method allows for identifying AREC areas that consist of one or more administrative areas, a limitation of the method still is that it divides the total region into several administrative areas. On the one hand, given that the method is especially developed for use on a low geographical scale, this is not much of a problem. On the other hand, the only alternative is to switch to distance-based measures, requiring much more data, which in practice are not always available.

The method to define AREC areas developed in this article provides a basis for future research. An analysis into the relationships between these areas and geographical characteristics, such as infrastructure availability and land value, can result in highly valuable insights for policy makers. In addition, it would be interesting to compare the employment development in these areas with the employment development in other areas in the region to see whether spatial concentration is important for firm and industry growth as well as the financial performance of co-located firms (although financial data on an establishment level are generally hard to collect). Since the method is relatively easy to apply and considers all intuitively important aspects of spatial concentration, these future research steps can serve as important input for policy decisions.

Acknowledgements

The authors are grateful to the Provincie Noord-Brabant for financial support, provision of data and helpful discussions.

Notes

1. Since this average is highly influenced by the number of areas in which the industry’s firms are located, this average has to be calculated only based on the areas in which there actually is employment in the
industry, which makes the method less dependent on the size of the industry and the number of areas defined.

2. The Nomenclature of Territorial Units for Statistics (NUTS) was established by Eurostat in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. NUTS 2 regions are identified as basic regions for the application of regional policies (Eurostat, 2011).

3. An LQ equal to 2 means that the share of industry employment is twice as large as the share of total employment in that particular area; all areas that are that specialized in a particular industry should be taken into account in the analysis.

4. It does not make sense to combine neighbouring areas if only one variable is used. Especially in the case of only using an absolute cut-off value, this would mean that all areas are combined into one area, since combining areas always results in more absolute employment. In the case of only using a relative cut-off value, combining neighbouring areas could in principle be done. However, also in this case combining neighbouring areas does not give additional value. If only the LQ is used as cut-off value, one area with only some employment and (almost) all of this employment in industry $i$ (“the shepherd in the desert”), can be combined with almost all neighbouring areas, since its LQ is that large that it will compensate for almost all low LQ values.

References


European Commission (2011) Promotion and Development of World-Class Clusters in Europe, Call for Proposals (Brussels: Enterprise and Industry Directorate-General). Available at http://ec.europa.eu/enterprise/newsroom/cf/itemdetail.cfm?item_id=4968&lang=en&tpa=0&displayType=fo&ref=newsbytheme%2Ecfrm%3Dlang%3Den%26displ...%3D0%26period%3Dlatest%26month%3D0%26page%3D1 (accessed 31 October 2012).


Appendix 1. Logistics SBI Codes

To classify establishments in different industries, the standard Dutch industry classification, the *Standaard BedrijfsIndeling* (SBI), is used. The SBI is developed by Statistics Netherlands and categorizes economic activities based on five digits. The first four digits correspond with the categorization of the European Union (NACE: *Nomenclature statistique des Activités économiques dans la Communauté Européenne*), with a small number of exceptions. The first two digits of the SBI and the NACE correspond to the categorization of the United Nations (ISIC: *International Standard Industrial Classification of All Economic Activities*).

Table A1 presents the SBI 2008 codes of the five industries used in this article. Since no SBI code exists for logistics, several codes are combined to define the logistics industry. In this article, the logistics industry is defined to consist of the following industries: wholesale trade, freight transport, cargo handling, storage and warehousing, and other supporting transport activities. From all establishments characterized as wholesale trade establishments based on the SBI code (46), the categories wholesale on a fee or contract basis (SBI = 461), wholesale of live animals (4623), and wholesale in computers, computer peripheral equipment and software (4651) are excluded. It is clear that indeed the establishments in the first two categories are no part of the logistics industry. In the last category mentioned, the wholesale in software is very dominant, definitely being no logistics, and hence, this category is also excluded from the analysis. In addition, to exclude wholesale trade establishments that are only responsible for the administrative part of trade and not the physical part, all wholesale trade establishments with less than 10 employees were deleted from the database. Administrative trade establishments are mostly relatively small and hence, this seems to be a valid method to exclude these establishments. Furthermore, all establishments in the above-described logistics categories with only one employee were excluded, since for these establishments it generally holds that the establishment’s address is equal to the owner’s address, which does not have anything to do with spatial concentration.

**Table A1.** SBI codes for the industries used in this article

<table>
<thead>
<tr>
<th>Name (based on ISIC)</th>
<th>SBI 2008 code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical production</td>
<td>C: 20</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>M: 72</td>
</tr>
<tr>
<td>Logistics</td>
<td></td>
</tr>
<tr>
<td>Wholesale trade and commission trade (except of motor vehicles and motorcycles)</td>
<td>G: 46</td>
</tr>
<tr>
<td>Freight transport via railways</td>
<td>H: 4920</td>
</tr>
<tr>
<td>Freight transport by road (except for removal transport)</td>
<td>H: 4921</td>
</tr>
<tr>
<td>Inland water ways freight transport</td>
<td>H: 50401−50403</td>
</tr>
<tr>
<td>Air freight transport</td>
<td>H: 5121</td>
</tr>
<tr>
<td>Storage and warehousing</td>
<td>H: 52101−52109</td>
</tr>
<tr>
<td>Other supporting transport activities</td>
<td>H: 52242, 52291−52292</td>
</tr>
<tr>
<td>Retail trade</td>
<td>G: 47</td>
</tr>
<tr>
<td>Construction</td>
<td>F</td>
</tr>
</tbody>
</table>
Appendix 2. Comparison of AREC Areas and Concentration Areas

Similar to Figure 5, Figures A1 and A2 compare the AREC areas identified for the chemical and R&D industry, respectively, with the areas that would have been identified based on only absolute employment concentration or relative employment concentration.

**Figure A1.** (a) Absolute chemical employment concentration. (b) Relative chemical employment concentration (LQ). (c) AREC areas for chemical production (also presented in Figure 4).
Figure A1. Continued.

Figure A2. (a) Absolute R&D employment concentration. (b) Relative R&D employment concentration (LQ). (c) AREC areas for R&D.
Figure A2. Continued.